#13

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	**Chris A. Hamilton**
Title of Invention	:	IMPROVED CONTROL OF VIDEOCONFERENCING USING ACTIVITY DETECTION
Application Filed	:	**June 6, 2000**
Application No.	:	**09/587,990**
Examiner	:	**Eng, George**
Art Unit	:	**2643**

Box Fee Amendment
Commissioner for Patents
Washington, DC 20231

AMENDMENT

Sir:

This is responsive to the Office Action dated June 7, 2002 in connection with the above-referenced patent application. A Petition for a three-month extension of time is enclosed herewith. A Terminal Disclaimer is also enclosed and is believed to overcome the obvious type double patenting and rejections, and withdrawal of such rejections is therefore requested at this time.

In parts 4 and 5 of the Action, the Examiner has rejected claim 15 over Zhou. The Examiner asserts that Zhou teaches a method for determining whether a conferee in a video conference is speaking by deciding whether "lip movements of said conferee are reasonably consistent with an audio signal from a conference station in which said conferee is located." Applicant respectfully disagrees with the Examiner's assertion that Zhou teaches a method of determining whether the conferee is speaking by comparing the lip movement with audio.

As explained at Col. 2 of Zhou, when the lips of a person in a video are moving, Zhou concludes that the conferee is speaking, and the audio signal is then encoded with greater accuracy. (Col. 2, lines 1-5). In a different embodiment, if the system detects lip movement

when there is audio, then the lip region may be encoded more accurately than the remainder of the video signal. (Col. 2, lines 22-30). Further, at Col. 18, lines 42-45, Zhou states that based upon a particular sound produced, the position of the mouth may be predicted, and the image enhanced.

Zhou therefore is concerned with accurately transmitting both voice and image data. When the lips are moving, Zhou assumes that the person is speaking and may enhance the audio signal from that conferee. No comparison of the audio and video is performed to determine that speech is occurring. Alternatively, when sound is detected, Zhou may enhance the image region around the conferee's lips. However, no place does Zhou disclose the step of correlating a specific audio signal to the specific video signal to determine if the two of them match so that the audio signal actually does represent human speech.

Indeed, Zhou is a specific teach away. More specifically, Zhou states in the abstract "A lip motion detection subroutine will detect the location and movement of the lips of a person present in a video scene in order to determine when a person is speaking and to encode the lip regions more accurately." Therefore, Zhou uses only video to determine that the person is speaking by lip movements, and not a match between the audio and the video, see if the sound coming out is reasonably consistent with the type of movement of the person's lips.

Similarly, at Col. 18 of Zhou, lines 39-49, Zhou discloses a system that analyzes the speech sound only in order to predict the position of the lips. Then, in response to the sound only, the conferee is presumed to be speaking, and the lip portion of the video may be enhanced. Thus, this embodiment merely discloses detecting when a person is speaking by using the audio signal only, and enhancing the video image based upon such detection. Once again, however, the determination as to whether or not the person is speaking is not based upon comparing the video image to the audio image in order to determine whether the audio is such that it is reasonably consistent with the video lip movements. Indeed, Zhou examines either the audio or the video to determine that speech is occurring, and then assumes such speech based upon only one signal, either the audio or the video. The fact that the other signal is enhanced in response to the detection of speech does not detract from fact that Zhou is merely teaching the use of either audio or video to determine when the conferee is speaking.

In contrast to Zhou, present claim 15 calls for the analysis of whether visual lip movements of a conferee are reasonably consistent with an audio signal from a conference station such that the combination of lip movement and audio signal indicates human speech. Applicant respectfully submits that Zhou does not disclose such a system, but only relies upon either audio or video to ascertain when speech is occurring, but not both.

Turning to the remainder of the Office Action, claims 5-14 are rejected as unpatentable over Ogata or Kamata in view of Zhou. Claims 5-14 all require, at one form or another, that both the lip movement of the video signal and the audio signal be utilized in a determination as to whether or not the party is speaking. It is respectfully submitted that none of the prior art discloses such an arrangement. None of the prior art discloses the critical and unique step of comparing the audio signal to the video signal to determine if the two are reasonably consistent with one another. Instead, Zhou simply shows the use of either audio or video to recognize the speech, and then using that recognition to enhance one of the signals. Therefore, reconsideration and allowance are respectfully requested.

Respectfully submitted,

KAPLAN & GILMAN, L.L.P.
900 Route 9 North
Woodbridge, New Jersey 07095
Telephone (732) 634-7634

DATED: November 22, 2002

Jeffrey B. Kaplan
(Reg. No. 34,356)